

An *LS-Decomposition* Approach for Robust Data Recovery in Wireless Sensor Networks

Xiao-Yang Liu, Xiaodong Wang, *Fellow, IEEE*, Linghe Kong, Meikang Qiu,
and Min-You Wu, *Senior Member, IEEE*

Abstract

Wireless sensor networks are widely adopted in military, civilian and commercial applications, which fuels an exponential explosion of sensory data. However, a major challenge to deploy effective sensing systems is the presence of *massive missing entries, measurement noise, and anomaly readings*. Existing works assume that sensory data matrices have low-rank structures. This does not hold in reality due to anomaly readings, causing serious performance degradation. In this paper, we introduce an *LS-Decomposition* approach for robust sensory data recovery, which decomposes a corrupted data matrix as the superposition of a low-rank matrix and a sparse anomaly matrix. First, we prove that LS-Decomposition solves a convex program with bounded approximation error. Second, using data sets from the IntelLab, GreenOrbs, and NBDC-CTD projects, we find that sensory data matrices contain anomaly readings. Third, we propose an accelerated proximal gradient algorithm and prove that it approximates the optimal solution with convergence rate $O(1/k^2)$ (k is the number of iterations). Evaluations on real data sets show that our scheme achieves recovery error $\leq 5\%$ for sampling rate $\geq 50\%$ and almost exact recovery for sampling rate $\geq 60\%$, while state-of-the-art methods have error $10\% \sim 15\%$ at sampling rate 90% .

Index Terms

Wireless sensor networks, robust data recovery, LS-decomposition

X. Liu and X. Wang are with the Department of Electrical Engineering, Columbia University.

X. Liu and L. Kong are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University.

M. Qiu is with the Department of Computer Science, Pace University.

I. INTRODUCTION

Wireless sensor networks (WSNs) [1] are constantly generating an enormous amount of rich and diverse information. WSNs are widely adopted in military, civilian and commercial applications, such as intrusion detection in battlefields [2], search and rescue systems [3], infrastructure monitoring [4], environment monitoring etc. The increasing number of big data sources have fueled an exponential explosion of sensory data [5]. Utilizing such a huge amount of sensory data for information substraction and interpretation, we are able to quantitatively understand physical phenomena and perform actively control over cyber-physical systems (CPS).

However, a major challenge to deploy effective sensor systems is the presence of *massive missing entries, measurement noise, and anomaly readings*. Missing entries will detriment the usability and reliability of the sensory data sets, while measurement noise and anomaly readings will cause erroneous conclusions and decisions. The data loss rate in real-world projects can be as high as 5% \sim 95%, as shown in Fig. 4 Section V, due to unreliable wireless communications such as poor link quality and packet collision. Moreover, measurement noise and anomaly readings are ubiquitous in real-world projects mainly due to: (1) commodity sensors have low accuracy; (2) some sensor nodes have malfunctions; and (3) the dynamic surroundings may cause disturbances. Therefore, it is necessary to distinguish environmental data from anomalies readings and measurement noise in a robust and accurate way.

Network data are naturally represented as a matrix in which rows denote nodes and columns denote time slots. Therefore, many recent works apply matrix completion techniques to perform data recovery [6–9] and data collection [10, 11] in WSNs. Compression is performed in the data collection phase to reduce communication burden, while matrix completion is applied at the sink node to recover the raw sensory data. However, the standard matrix completion [12–14] experiences serious performance degradation with the presence of a small portion of anomalies. Fig. 1 shows a simple example, a 100×100 matrix with rank $r = 5$ taking values in $[0, 100]$, can be exactly recovered using 30% uniformly random selected samples [14]. However, if we add anomalies with value = 100, two difference ratios 0.5% and 3%, the recovery errors (ℓ_2 -norm) increase significantly.

The performance degradation comes from the fact that: even a small portion of anomaly readings can break down the low-rank structure. As shown in Fig. 2, adding 0.5% anomalies into the above rank $r = 5$ matrix, the energy captured in the top 10 singular values is 80%, while it is only 53% when 3% anomalies are added. In Section V, using data sets collected from the IntelLab [15], GreenOrbs [16], and NBDC-CTD [17] projects, we observe that anomaly readings are common and ubiquitous. Therefore, we

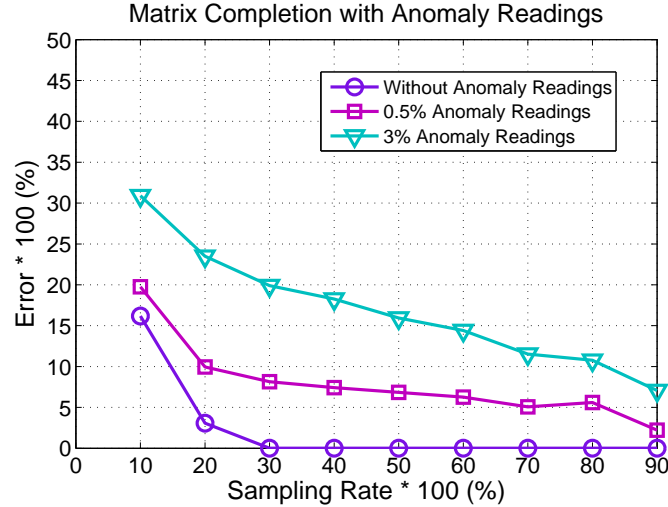


Fig. 1. Performance of matrix completion: exactly low-rank matrix, compared with matrices with 0.5% and 3% anomalies.

find that existing works hastily assume low-rank structures.

In this paper, we introduce an *LS-Decomposition* approach for sensory data recovery, which can deal with massive data loss and is robust to both measurement noise and anomaly readings. By modeling measurement noise as small entry-wise noise, and anomaly reading as gross sparse error, LS-Decomposition decomposes a sensory data matrix into a low-rank matrix and a sparse anomaly matrix. Firstly, we present observations on real data sets from the IntelLab [15], GreenOrbs [16], and NBDC-CTD [17] projects, showing that (1) the original data matrices have massive missing entries and (2) each data matrix is the superposition of a low-rank matrix and a sparse anomaly matrix. Secondly, we formulate the robust data recovery problem as an optimization problem, coined as *LS-Decomposition*, and prove that solving a convex program achieves bounded approximation. Thirdly, we propose an accelerated proximal gradient algorithm for LS-Decomposition, theoretical results guarantee that our algorithm approximates the optimal solution with convergence rate $O(1/k^2)$ (k is the number of iterations). Finally, evaluations on real data sets show that our scheme achieves recovery error $\leq 5\%$ for sampling rate $\geq 50\%$ and almost exact recovery for sampling rate $\geq 60\%$, while state-of-the-art methods have error $10\% \sim 15\%$ at sampling rate 90% .

The reminder of the paper is organized as follows. Section II discusses related works. System models and problem formulation are given in Section III. We present conditions and theoretic guarantees for optimal LS-Decomposition in Section IV. Observations are presented in Section V. Our algorithm is in Section VI, while Section VII describes performance evaluation. We conclude in Section VIII. The appendix provides detailed proofs.

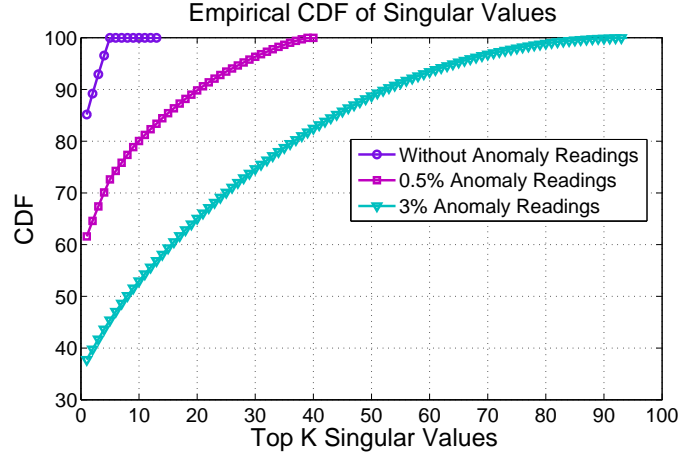


Fig. 2. Empirical CDF of energy captured by the top K singular values: exactly low-rank matrix, compared with matrix with 0.5% and 3% anomalies.

II. RELATED WORK

We first describe the applications of the standard matrix completion in wireless sensor networks. Then, we discuss several variants of matrix completion.

Data loss is revealed to be ubiquitous and unavoidable in sensor networks [6]. In order to improve the reliability and useability of decisions draw from such incomplete sensory data, a data recovery process is needed. Since the raw sensory data are redundant [18], it is able to estimate the original environment data from partial observations. In [6], authors show that data matrices of temperature, humidity and light are low-rank and have high spatiotemporal correlations, while similar empirical observations are also presented in [11]. To recover effectively, [6, 9] apply data preprocess on the raw data sets to exclude anomaly readings, and then matrix completion to perform data recovery. However, there are no good criteria for identifying anomaly readings while [6, 9] are based on researchers' experience. Furthermore, the inter-correlation among multiple attributes are exploited to improve the recovery accuracy [7]. Matrix completion is also suitable for seismic data recovery [8].

Besides data recovery, matrix completion is also applied to improve data collection [10, 11] in WSNs. Raw-data collection is rather inefficient [19], since WSNs are typically composed of hundreds to thousands of sensor nodes generating tremendous amount of sensory readings. As the packet loss problem and the hot spot problem surface, this approach will lead to a large number of retransmissions in real-world situations and node failures as batteries run out. Therefore, researchers apply matrix completion to reduce global traffic [10]. Compression is performed in the data collection phase to reduce communication burden, while matrix completion is applied at the sink node to recover the raw sensory data. Furthermore, the

authors of [11] propose an online data gathering scheme for weather data.

A natural technique to improve the recovery performance of the standard matrix completion is to add a smooth regulation term [6, 7, 9, 11]. This intuition is inspired by the fact that almost all physical conditions are smooth fields, i.e., the physical conditions are continuous without sudden changes. However, all the above schemes are based on a critical assumption that the matrix of interest exhibits a low-rank structure, which does not hold in reality. As shown in Fig. 2, 4 and 5, real-world data matrices violate such assumption due to the existence of a small portion of anomaly readings. Therefore, the standard matrix completion [12, 13] and the smoothed counterpart experience serious performance degradation in practical scenarios.

Noticing the ubiquitousness of anomalies in practical scenarios, researchers propose to decompose the data matrix into a low-rank matrix and a sparse anomaly matrix [20–22] and prove its universal applicability. The authors prove that it is possible to exactly separate those two components. These works inspire us in proving that our problem has bounded approximation, however, our work essentially differs from theirs in terms of models and goals: (1) no measurement noise is allowed in their model. They consider an *exactly* low-rank matrix, while we deal with an *approximately* low-rank matrix which is the case for sensory data matrices; (2) full observations of all entries are required. They target at the possibility of *separating* two matrices while we aim to *recover* the low-rank matrix.

III. SYSTEM MODEL AND PROBLEM STATEMENT

A. Notations

We go over the notations and preliminaries. Throughout the paper, N denotes the number of nodes in the wireless sensor network, T denotes the number of slots in the sensing period of interest, $[N]$ denotes the set $\{1, 2, \dots, N\}$, $R \in \mathbb{R}^{N \times T}$ denotes the data matrix of interest, $L \in \mathbb{R}^{N \times T}$ denotes the low-rank matrix of interest, and $S \in \mathbb{R}^{N \times T}$ denotes the anomaly matrix of interest. Let $\Omega \subset [N] \times [T]$ denote the support set of S (i.e., the index set of the nonzero entries), $\mathcal{O} \subset [N] \times [T]$ denote the index set of the observed entries. R^* denotes the transpose of R . The operator $\mathcal{P}_{\mathcal{O}}(R)$ projects matrix R onto its index set \mathcal{O} , i.e., the (n, t) -th entry of $\mathcal{P}_{\mathcal{O}}(R)$ is equal to R_{nt} if $(n, t) \in \mathcal{O}$ and zero otherwise.

The ℓ_0 -norm is defined as $\|S\|_0 \doteq |\text{supp}(S)|$, where $\text{supp}(S) \doteq \{(n, t) : S_{nt} \neq 0\}$ denotes the support set of S and $|\text{supp}(S)|$ denotes the cardinality of $\text{supp}(S)$. The ℓ_1 -norm is defined as $\|S\|_1 \doteq \sum_{n=1}^N \sum_{t=1}^T \text{abs}(S_{nt})$ where $\text{abs}(S_{nt})$ denotes the absolute value of S_{nt} . The ℓ_2 -norm of vector $A \in \mathbb{R}^m$ is defined as $\|A\|_2 = \sqrt{\sum_{i=1}^m A_i^2}$. The Frobenius norm of matrix R is defined as $\|R\|_F \doteq \sqrt{\sum_{n=1}^N \sum_{t=1}^T R_{nt}^2}$. Let $L = U\Sigma V^* = \sum_{i=1}^r \sigma_i u_i v_i^*$ denote the singular value decomposition (SVD)

of L , where r is the rank, $\sigma_1 \geq \dots \geq \sigma_r > 0$ are the singular values, and $U = [u_1, \dots, u_r]$, $V = [v_1, \dots, v_r]$ are the matrices of left- and right-singular vectors, respectively. The nuclear norm of L is defined as $\|L\|_* = \sum_{i=1}^r \sigma_i$. Given a matrix pair $X = \langle L, S \rangle$, let $\|X\|_F \triangleq (\|L\|_F^2 + \|S\|_F^2)^{1/2}$.

More notations are needed in our proofs as in [20–22]. Let \mathcal{T} denote the subspace generated by matrices with the same column space or row space of L :

$$\mathcal{T} = \{UQ^* + RV^*|Q, R \in \mathbb{R}^{n \times r}\} \subseteq \mathbb{R}^{n \times n}, \quad (1)$$

and $\mathcal{P}_{\mathcal{T}}$ be the projection operator onto the subspace \mathcal{T} . Define the projection operator $\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_{\Omega} : (L, S) \mapsto (\mathcal{P}_{\mathcal{T}}L, \mathcal{P}_{\Omega}S)$. Define the subspace $\Gamma \triangleq \{(Q, Q)|Q \in \mathbb{R}^{n \times n}\}$ and $\Gamma^{\perp} \triangleq \{(Q, -Q)|Q \in \mathbb{R}^{n \times n}\}$, and let \mathcal{P}_{Γ} and $\mathcal{P}_{\Gamma^{\perp}}$ denote their respective projection operators.

B. Network Model

We consider a wireless sensor network consisting of N sensor nodes and a sink. Sensor nodes are scattered in a target field to monitor physical conditions such as temperature, humidity, illumination, gas concentration, magnetic strength, etc. They report sensory readings to the sink periodically over a given time span.

The monitoring period is evenly divided into T time slots, denoted as $\{0, 1, \dots, t, \dots, T-1\}$. Each sensor node generates a record in each slot. A record at a sensor node includes the sensor readings, node ID, time stamp, and location (longitude and latitude). Its format is:

Record:	sensor readings	node ID	time stamp	location
---------	-----------------	---------	------------	----------

Let $R_{i,t}$ denote the sensory reading of the i -th node at slot t . For each physical condition, the sensor readings generated by all sensor nodes can be represented by a matrix $R \in \mathbb{R}^{N \times T}$, as follows:

$$R = \begin{bmatrix} R_{0,0} & \dots & R_{0,t} & \dots & R_{0,T-1} \\ R_{1,0} & \dots & R_{1,t} & \dots & R_{1,T-1} \\ \dots & \dots & \dots & \dots & \dots \\ R_{N-1,0} & \dots & R_{N-1,t} & \dots & R_{N-1,T-1} \end{bmatrix} \quad (2)$$

where rows denote nodes, and columns denote time slots.

C. Measurement Model

It is widely believed that sensory data exhibit strong spatio-temporal correlation [18?]. Compressive sensing theory [12, 13] introduces the general notion of “low-rank” to model this characteristic. In this

paper, we assume that the actual sensory data of a physical condition is an *approximately* low-rank matrix $L \in \mathbb{R}^{N \times T}$ as in [6–11], which means that most of its energy is captured by its rank- r approximation. The observed sensory reading matrix R was generated by corrupting the entries of L with measurement noise and anomaly readings. We model measurement noise as additive small entry-wise noise, and anomaly readings as additive gross sparse errors, respectively, i.e., $Z \in \mathbb{R}^{N \times T}$ with $\|Z\|_F \leq \delta$ for some $\delta > 0$, and $S \in \mathbb{R}^{N \times T}$ with $\|S\|_0 \leq k$. It is reasonable to assume that the number of anomaly readings is relatively small, compared with the size of the sensory data matrix, i.e., $k \ll NT$. Let $\Omega \subset [N] \times [T]$ denote the support set of S , i.e., the index set of the nonzero entries. We have the following measurement model:

$$R = L + S + Z. \quad (3)$$

D. Problem Statement

Sensor nodes cooperate with each other to transmit packets back to the sink via multi-hop wireless communication. Two major factors lead to massive missing entries: (1) in each hop, poor link quality results in decoding failures at the receiver node; (2) along the multi-hop path, packet collisions are unavoidable since wireless communication utilizes unregulated media access.

Therefore, the sink receives an incomplete measurement matrix M . The data loss rate can be as high as 5% \sim 95% as shown in Fig. 4 Section VI. Suppose the collected entries are indicated by the set $\mathcal{O} \subset [N] \times [T]$ and \mathcal{O} has size m . We assume that the missing entries are randomly distributed, either uniformly random or non-uniform random. Please refer to [20] for more mathematic details about allowed distributions. The data collection process can be represented as:

$$M = \mathcal{P}_{\mathcal{O}}(R), \quad (4)$$

where the operator $\mathcal{P}_{\mathcal{O}}(X)$ projects X onto the index set \mathcal{O} .

To estimate the low-rank matrix L from the partial observations M , a direct conceptual solution is: to seek a L with the lowest-rank that could have generated the data matrix R , subject to the constraints that the gross errors are sparse and the entry-wise errors are small. We formulate the *LS-Decomposition* problem as follows:

Problem 1. (LSD) Assuming $R = L + S + Z$ and given its partial observation $M = \mathcal{P}_{\mathcal{O}}(R)$ on index set \mathcal{O} , where L, S, Z are unknown, but L is known to be low-rank, S is known to be sparse, and $\|Z\|_F \leq \delta$

for some $\delta > 0$, recover L . The Lagrangian formulation is:

$$\begin{aligned} \langle \hat{L}, \hat{S} \rangle &= \min_{\langle L, S \rangle} \text{rank}(L) + \lambda \|S\|_0, \\ \text{s.t. } &\|\mathcal{P}_{\mathcal{O}}(R - L - S)\|_F \leq \delta, \end{aligned} \quad (5)$$

where we choose $\lambda = 1/\sqrt{n}$, and \hat{L}, \hat{S} are estimates of L, S , respectively.

IV. EXISTING RESULTS FOR LS-SEPARATION

A special case of the LSD problem (5) is when we have full observation of R , then (5) is reduced to the following *LS-Separation* problem studied in [20–22]. In this section, we summarize the conditions for unique solution and existing results.

Problem 2. (LSS) Assuming $R = L + S + Z$ and given full observation of R as $\mathcal{O} = [N] \times [T]$, recover the components L and S . The Lagrangian formulation is:

$$\begin{aligned} \langle \hat{L}, \hat{S} \rangle &= \min_{\langle L, S \rangle} \text{rank}(L) + \lambda \|S\|_0, \\ \text{s.t. } &\|R - L - S\|_F \leq \delta. \end{aligned} \quad (6)$$

The LSS problem (6) is a highly nonconvex optimization problem and no efficient solution is known, since (6) subsumes both the low-rank matrix completion problem and the ℓ_0 -minimization problem. Both of them are NP-hard and hard to approximate [29]. To obtain a tractable optimization problem, one can relax the LSS problem (6) by replacing the ℓ_0 -norm with the ℓ_1 -norm and the rank function with the nuclear norm as in [12, 20–22, 24, 26], yielding the following problem:

Problem 3. The convex surrogate of the LSS problem (6) is:

$$\begin{aligned} \langle \hat{L}, \hat{S} \rangle &= \min_{\langle L, S \rangle} \|L\|_* + \lambda \|S\|_1, \\ \text{s.t. } &\|R - L - S\|_F \leq \delta \end{aligned} \quad (7)$$

A. Conditions for Unique Solution

Before analyzing the optimality of problem (7), one should answer the following a basic question first: When is LS-Separation possible? At first sight, one may think that the objective of problem (5) is impractical. For instance, suppose the matrix R is equal to $e_1 e_1^*$ (e_1 is a canonical basis vectors, the resulting matrix $e_1 e_1^*$ has a one in the top left corner and zeros everywhere else), then since R is both sparse and low-rank, how can we decide whether it is low-rank or sparse? To make problem (5) meaningful, we need to impose that the low-rank component L is not sparse, and the sparse component S is not low-rank.

1) *The Low-Rank Matrix is Not Sparse:* Let $L = U\Sigma V^* = \sum_{i=1}^r \sigma_i u_i v_i^*$ denote the singular value decomposition (SVD) of $L \in \mathbb{R}^{N \times T}$, where r is the rank, $\sigma_1, \dots, \sigma_r$ are the singular values, and $U = [u_1, \dots, u_r], V = [v_1, \dots, v_r]$ are the matrices of left- and right-singular vectors, respectively.

The incoherence condition introduced in [12, 13] asserts that for small values of parameter μ , the singular vectors are reasonably spread out, namely, the low-rank matrix L is not sparse. It states that:

$$\max_i \|U^* e_i\|^2 \leq \frac{\mu r}{N}, \quad \max_i \|V^* e_i\|^2 \leq \frac{\mu r}{T}, \quad \|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{NT}}, \quad (8)$$

where e_i 's are the canonical basis vectors.

2) *The Sparse Matrix is Not Low-Rank:* Another identifiability issue arises if the sparse matrix $S \in \mathbb{R}^{N \times T}$ has low-rank. For instance, all the nonzeros entries of S lie in a column or in a few columns, suppose that the columns of S happens to be the opposite of those of L , then it is clear that we would not be able to recover L and S by any method whatsoever since $R = L + S$ would have a column space equal to, or included in that of L . To avoid such pathological cases, we assume that the support of sparse component S is selected uniformly at random among all subsets of size k as in [20–22].

B. Existing Main Results

The main result of [22] is that under the above conditions, (7) gives a stable estimate of L and S .

Lemma 1. [22] *Suppose that L obeys (8) and the support of S is uniformly distributed. Then if $\text{rank}(L) \leq \rho_r n \mu^{-1} (\log n)^{-2}$ and $|\Omega| \leq \rho_s n^2$ with $\rho_r, \rho_s > 0$ being sufficiently small numerical constants. Assume $\|\mathcal{P}_\Omega \mathcal{P}_T\| \leq 1/2, \lambda \leq 1/2$, let $\hat{X} = \langle \hat{L}, \hat{S} \rangle$ be the solution to (7). Then with high probability, \hat{X} satisfies:*

$$\|\hat{X} - X\|_F = \|\hat{L} + \hat{S} - L - S\|_F \leq (8\sqrt{5}n + \sqrt{2})\delta, \quad (9)$$

where $n = \max(N, T)$.

We can see that (7) is simultaneously stable to small entry-wise noise and robust to gross sparse anomalies. Note that the error bound for each entry is proportional to the noise level δ . As for “with high probability”, there is no exact result while [20–22] proved that the probability is at least $1 - cn^{-10}$ where c is a numerical constant.

V. CONDITIONS AND PERFORMANCE GUARANTEES FOR LS-DECOMPOSITION

A. Problem Analysis

Following the framework in Section IV, we relax problem (5) by replacing the ℓ_0 -norm with the ℓ_1 -norm and the rank function with the nuclear norm. Then, we provide a condition for unique solution and

the corresponding theoretical guarantees.

Problem 4. *The convex surrogate of the LSD problem (5) is:*

$$\begin{aligned} \langle \hat{L}, \hat{S} \rangle &= \min_{\langle L, S \rangle} \|L\|_* + \lambda \|S\|_1, \\ \text{s.t. } &\|\mathcal{P}_{\mathcal{O}}(R - L - S)\|_F \leq \delta \end{aligned} \quad (10)$$

B. Partial Observation

However, those two conditions in Section IV do not suffice to guarantee unique solution for problem (5). For example, all $\{\langle L, S \rangle, \langle L, S_1 \rangle, \dots, \langle L, S_n \rangle\}$ are optimal solutions as long as $\mathcal{P}_{\mathcal{O}}(S_1) = \dots \mathcal{P}_{\mathcal{O}}(S_n) = \mathcal{P}_{\mathcal{O}}(S)$. To avoid such predicament, we set the following artificial setting for (5) and (7) throughout the paper:

$$\|\mathcal{P}_{\mathcal{O}^\perp}(\hat{S})\|_F = \|\mathcal{P}_{\mathcal{O}^\perp}(S_1)\|_F = \dots = \|\mathcal{P}_{\mathcal{O}^\perp}(S)\|_F = 0. \quad (11)$$

This is quite reasonable because our aim is to recover the low-rank matrix L and we do not want to recover the anomaly matrix.

Remark 1. *Partial observation of R provides partial recovery of its sparse component S , i.e., only the corresponding subset of entries are observable.*

C. Theoretic Guarantees

Under the above conditions, we have the following theoretical results: 1) problem (5) is possible by solving the convex program (10), and 2) the precise closed form of the approximation bound. The detailed proofs are given in the appendix.

Theorem 1. *Suppose that L obeys (8) and the support of S is uniformly distributed. Then if $\text{rank}(L) \leq \rho_r n \mu^{-1} (\log n)^{-2}$ and $|\Omega| \leq \rho_s n^2$ with $\rho_r, \rho_s > 0$ being sufficiently small numerical constants. Assume $\|\mathcal{P}_{\Omega} \mathcal{P}_{\mathcal{T}}\| \leq 1/2, \lambda \leq 1/2$, let $\hat{X} = \langle \hat{L}, \hat{S} \rangle$ be the solution to (10). Then with high probability, \hat{X} satisfies:*

$$\|\hat{X} - X\|_F = \|\hat{L} + \hat{S} - L - S\|_F \leq (8n\sqrt{40n + 40n/p + 5} + \sqrt{2})\delta, \quad (12)$$

where $n = \max(N, T)$.

TABLE I
DATA SETS FOR THE COMPRESSIBILITY CHARACTERIZATION.

Data Sets	Environment	Nodes	Time period	Resolution	Physical conditions	Size
IntelLab	Indoor	54	Feb.28 ~ Apr.5, 2004	30 seconds	Temperature, light, humidity	54×500
GreenOrbs	Forest	326	Aug.03 ~ 05, 2011	60 seconds	Temperature, light, humidity	326×750
NBDC CTD	Ocean	216	Oct.26 ~ 28, 2012	10 minutes	Temperature, salt, conductivity	216×300

VI. REVEALING ANOMALY READINGS IN REAL-WORLD WIRELESS SENSOR NETWORKS

A. Data Sets

Table I lists the sensory data matrices used in this paper. These data sets are collected by the IntelLab [15], GreenOrbs [16], and NBDC-CTD [17] projects, which are deployed in indoor, mountain and ocean environments, respectively.

IntelLab [15]: is deploy in the Intel Berkeley Research lab from Feb. 28th to Apr. 5th, 2004. There are 54 Mica2 nodes places in a $40\text{m} \times 30\text{m}$ room. The nodes are set to send back a packet every 30 seconds. This data set includes in total 2,313,682 data packets.

GreenOrbs [16]: is deployed on the Tianmu Mountain, Zhejiang Province, China. In total, there are 326 nodes deployed. Nodes are set to transmitting packets back to the sink node every minute. We use the data collected in Aug. 3rd ~ 5th, 2011, including 308,928 data packets.

NBDC CTD [17]: by the National Oceanic and Atmospheric Administration's (NOAA) National Data Bouy Center (NDBC). CTD (Conductivity, Temperature, and Depth) is a shipboard device consisting of many small probes. Nodes are set to report every 10 minutes. We use 8,107 data packets collected during Oct. 26th ~ 28th, 2012.

B. Data Loss Rate

Due to unreliable wireless communication and packet collision, the raw data matrices at the sink node are incomplete. To quantify the data loss problem, we define a loss rate for each node as $r_i = \mathcal{L}_i/T$, $i \in \{1, 2, \dots, n\}$, where \mathcal{L}_i is the number of lost readings of the i -th sensor node. The empirical cumulative distribution function of data loss rates is shown in Fig. 4. The data loss rates in real-world sensor networks are quite high, being $5\% \sim 95\%$.

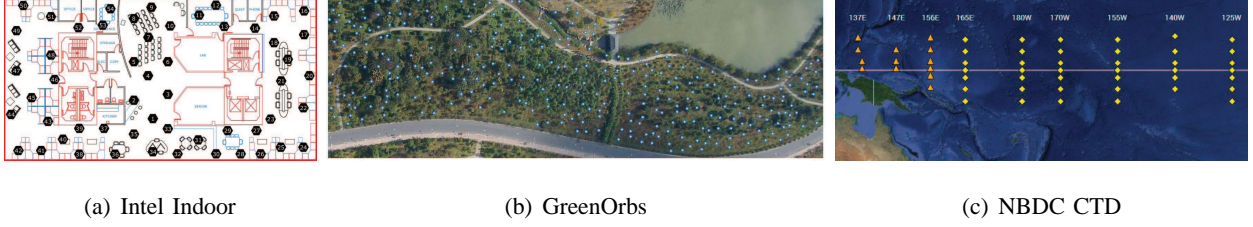


Fig. 3. Overview of the sensor network deployment for the Intel Indoor, GreenOrbs and NBDC CTD projects.

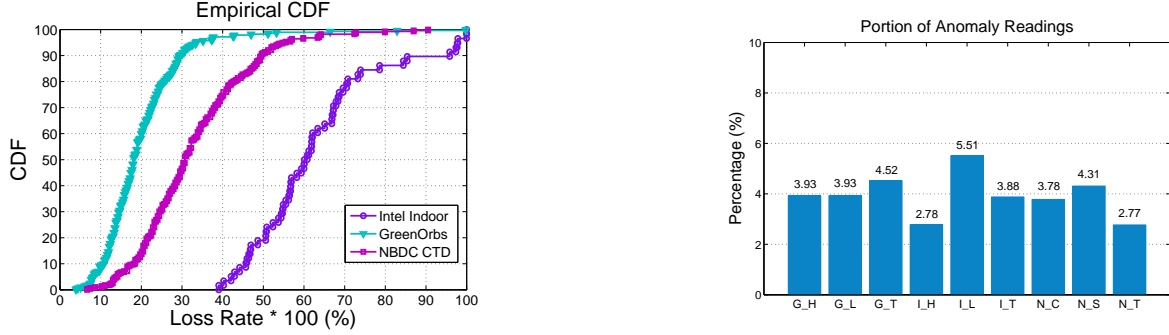


Fig. 4. (Left) The data loss rates of the IntelLab, GreenOrbs, and NBDC-CTD projects, deployed in indoor, mountain, and ocean environment, respectively. (Right) The percentage of anomaly readings, determined by the LS-Separation [22]. G,I,N are short for GreenOrbs, IntelLab, and NBDC CTD, while H, L, T, C, S are short for humidity, light, temperature, conductivity, and salt.

C. Low-Rank Property with Anomaly Readings

First, we construct full data matrices listed in Table. I, serving as the ground truth for observations and evaluations. Set a time window which is ten times the monitoring resolution for each data set. For each time slot, we randomly select one of the observed readings.

For each raw data matrix, we apply singular value decomposition (SVD) to examine if it has a good low-rank approximation. The metric we use is the fraction of total energy captured by the top portion of singular values. For example, the top 10% portion of singular values for the IntelLab matrices means

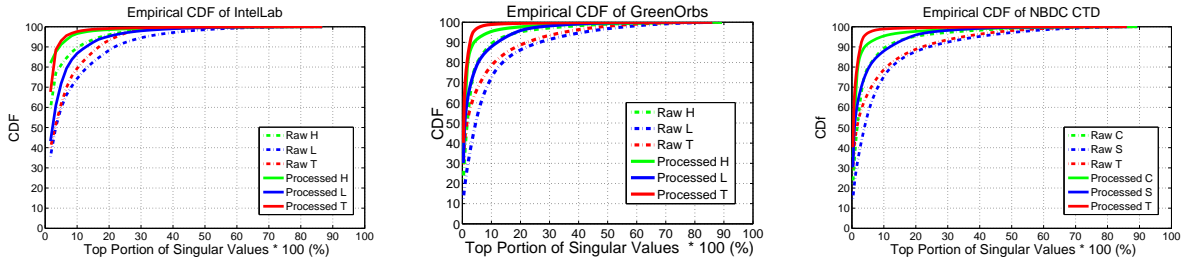


Fig. 5. The CDF of Singular Values for IntelLab, GreenOrbs, and NBDC CTD: svd of the raw data matrix, compared with processed data matrix. H, L, T, C, S are short for humidity, light, temperature, conductivity, and salt.

the largest 5 singular values. We calculate $\left(\sum_{i=1}^5 \sigma_i\right) / \left(\sum_{i=1}^{54} \sigma_i\right)$, where σ_i is the i -th largest singular value and $\left(\sum_{i=1}^{54} \sigma_i\right)$ measures the total energy of the data matrix. Note that $1 - \left(\sum_{i=1}^5 \sigma_i\right) / \left(\sum_{i=1}^{54} \sigma_i\right)$ is the relative approximation error of the best rank-5 approximation with respect to the Frobenius norm.

Next we process the raw data matrices with *LS-separation* [20][14], which identifies the anomaly readings and returns a low-rank matrix. Then, we apply SVD on the processed low-rank matrices in a similar way.

Fig. 4 shows the ratios of anomaly reading for each data matrix, ranging from 2.77% to 5.51%. Fig. 5 shows the CDFs of singular values for the raw data matrices and the processed data matrices. We can see that by excluding the anomalies, the processed data matrices has strengthened low-rank structures. For the IntelLab project, 18% singular values capture 90% energy of the raw temperature matrix while the ratio is reduced to 5%; for light, 22% is reduced to 13%; and for humidity, 10% is reduced to 5%. Similar observations hold for the GreenOrbs and NBDC CTD projects.

VII. ACCELERATED PROXIMAL GRADIENT APPROACH

A. General Accelerated Proximal Gradient Algorithm

Accelerated Proximal Gradient algorithms (APGs) [27], with the general flow shown in Table II, solve problems with the following general form:

$$\min_{X \in \mathcal{H}} F(X) \triangleq \mu g(X) + f(X), \quad (13)$$

where \mathcal{H} is a real Hilbert space with a norm $\|\cdot\|$, g is a continuous convex function, f is convex and smooth with Lipschitz continuous gradient: $\|\nabla f(X_1) - \nabla f(X_2)\| \leq L_f \|X_1 - X_2\|$ where ∇f is the Fréchet derivative of f .

APGs repeatedly minimize a proximal operator $Q(X, Y)$ to $F(X)$, defined as:

$$Q(X, Y) \triangleq f(Y) + \langle \nabla f(Y), X - Y \rangle + \frac{L_f}{2} \|X - Y\|^2 + \mu g(X).$$

Since $f(X)$ is convex, for any Y , $Q(X, Y)$ upper bounds $F(X)$. If we define $G \triangleq Y - \frac{1}{L_f} \nabla f(Y)$, with simple transform and ignore the constant terms, then we have:

$$\arg \min_X Q(X, Y) = \arg \min_X \mu g(X) + \frac{L_f}{2} \|X - G\|^2. \quad (14)$$

We repeatedly set $X_{k+1} = \arg \min_X Q(X, Y_k)$, as in Table II. It is proved [28] that setting $Y_k = X_k + \frac{t_{k-1}-1}{t_k}(X_k - X_{k-1})$ if $t_{k+1}^2 - t_{k+1} \leq t_k^2$ results in a convergence rate $O(1/k^2)$.

TABLE II
GENERAL PROXIMAL GRADIENT ALGORITHM

```

1: while not converged do
2:    $Y_k \leftarrow X_k + \frac{t_{k-1}-1}{t_k}(X_k - X_{k-1});$ 
3:    $G \leftarrow Y_k - \frac{1}{L_f}\nabla f(Y_k);$ 
4:    $X_{k+1} \leftarrow \arg \min_X \mu g(X) + \frac{L_f}{2}\|X - G_k\|^2;$ 
5:    $t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}, k \leftarrow k+1;$ 
6: end while

```

TABLE III
ACCELERATED PROXIMAL GRADIENT ALGORITHM

```

Input:  $M, \Omega, \lambda.$ 
1:  $L_0, L_{-1} \leftarrow 0; S_0, S_{-1} \leftarrow 1; t_0, t_{-1} \leftarrow 1; \bar{\mu} \leftarrow \delta\mu_0; k=0.$ 
2: while not converged do
3:    $Y_k^L \leftarrow L_k + \frac{t_{k-1}-1}{t_k}(L_k - L_{k-1}).$ 
4:    $Y_k^S \leftarrow S_k + \frac{t_{k-1}-1}{t_k}(S_k - S_{k-1}).$ 
5:    $G_k^L \leftarrow Y_k^L - \frac{1}{2}\mathcal{P}_O(Y_k^L + Y_k^S - R).$ 
6:    $(U, S, V) \leftarrow \text{svd}(G_k^L), L_k = US_{\frac{\mu_k}{2}}[S]V^*.$ 
7:    $G_k^S \leftarrow Y_k^S - \frac{1}{2}\mathcal{P}_O(Y_k^L + Y_k^S - R).$ 
8:    $S_{k+1} \leftarrow \mathcal{S}_{\frac{\lambda\mu}{2}}[G_k^S].$ 
9:    $t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}, \mu_{k+1} \leftarrow \max(\mu_k, \bar{\mu}), k \leftarrow k+1.$ 
10: end while
Output:  $\hat{L} \leftarrow L_k, \hat{S} \leftarrow S_k.$ 

```

B. Robust Data Recovery by Accelerated Proximal Gradient

Consider the augmented Lagrangian function of (10),

$$\arg \min_X F(X) \triangleq \mu\|L\|_* + \mu\lambda\|S\|_1 + \frac{1}{2}\|\mathcal{P}_O(R - L - S)\|_F^2. \quad (15)$$

Different from the conventional APGs, \mathcal{H} is the space of same-sized matrices endowed with the Frobenius norm $\|\cdot\|_F$, our iterates X_k are ordered pairs $(L_k, S_k) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, $g(X_k) = \mu\|L_k\|_* + \mu\lambda\|S_k\|_1$, and the Lipschitz constant $L_f = 2$.

Write $G_k \triangleq (G_k^L, G_k^S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, and let USV^* be the singular value decomposition of G_k^L . Following the framework of APGs, we have:

$$L_{k+1} = US_{\frac{\mu}{2}}[S]V^*, \quad S_{k+1} = \mathcal{S}_{\frac{\lambda\mu}{2}}[G_k^S], \quad (16)$$

where $\mathcal{S}[\cdot]$ is an element-wise threshold operator, defined as:

$$\mathcal{S}_\epsilon[x] = \begin{cases} x - \epsilon, & \text{if } x > \epsilon \\ x + \epsilon, & \text{if } x < -\epsilon \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

Therefore, we construct our accelerated proximal algorithm as in Table III. Please refer to the appendix for the proofs of the following two theorems.

Theorem 2. Let $F(X) \triangleq F(L, S) \triangleq \mu\|L\|_* + \mu\lambda\|S\|_1 + \frac{1}{2}\|\mathcal{P}_O(R - L - S)\|_F^2$. Then, for all $k \geq 1$, our algorithm achieves the following converge rate:

$$F(X_k) - F(X^*) \leq \frac{4\|X_1 - X^*\|_F^2}{k^2}, \quad (18)$$

where X^* is optimal solution to (10).

Lemma 2. Let $F(X) \triangleq F(L, S) \triangleq \mu\|L\|_* + \mu\lambda\|S\|_1 + \frac{1}{2}\|\mathcal{P}_O(R - L - S)\|_F^2$. Our algorithm converges to its global optimal. To achieve $F(X_k) - F(X^*) \leq \epsilon$, our algorithm requires $k = O(1/\sqrt{\epsilon})$ iterations.

VIII. EVALUATION

A. Evaluation Methodology

Our evaluations are based on the real-world data sets from the IntelLab [15], GreenOrbs [16], and NBDC-CTD [17] projects. The sensors in each project measure three types of physical conditions. Given the ground truth data described in Table I, we randomly drop entries and then compare the recovered data matrices with the ground truth. Under different data loss rate, we measure the recovery errors in terms of *Normalized Square Error (NSE)*, defined as following:

$$NSE = \frac{\sum_{i=1}^N \sum_{j=1}^T (\hat{L}_{ij} - L_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^T L_{ij}^2}, \quad (19)$$

where \hat{L} is the estimated low-rank matrix, L is the ground truth. *NSE* is widely used in data interpolation and parameter estimation.

We compare our LS-Decomposition with the standard matrix completion and the smooth-regulated matrix completion. To further understand the gap between LS-Decomposition and the optimal recovery

performance, we construct an oracle solution which gives us information about the support Ω of S and the row (column) space \mathcal{T} of L .

(1) Oracle Solution (OS) [22] serves as the optimal solution. Assume that we know the support Ω of S and the row and column spaces \mathcal{T} of L , which can be inferred by performing *LS-Separation* (problem P6 as done in Section V) on the ground truth data matrices. It estimates L and S as the solution L_{oracle} and S_{oracle} to the following least squares problem:

$$\begin{aligned} \langle L_{oracle}, S_{oracle} \rangle = \min_{\langle L, S \rangle} \|R - L - S\|_F, \\ s.t. \quad L \in \mathcal{T}, S \in \Omega. \end{aligned} \quad (20)$$

(2) Matrix Completion (MC) [12][13][14]: the standard matrix completion solves the following optimization problem:

$$\begin{aligned} \hat{L} = \min_L \|L\|_*, \\ s.t. \quad \|R - L\|_F \leq \delta. \end{aligned} \quad (21)$$

(3) SRMF [6]: it uses Sparsity Regularized Matrix Factorization that leverages both low-rank and spatio-temporal characteristics, as follows:

$$\begin{aligned} \hat{L} = \min_L \|L\|_* + \lambda \mathcal{S}(L), \\ s.t. \quad \|R - L\|_F \leq \delta, \end{aligned} \quad (22)$$

where $\mathcal{S}(\cdot)$ is the smooth term, and λ is the weight to balance the low-rank term and the smooth term, which is usually determined by experiments. We find $\lambda = 0.01$ gives the best performance in our simulations. $\mathcal{S}(L)$ is defined via the diversity of matrix horizontal and vertical difference:

$$\mathcal{S}(L) = \|\mathcal{D}_x(L)\|_F^2 + \|\mathcal{D}_y(L)\|_F^2, \quad (23)$$

where $\mathcal{D}_x(L)$ is an $N \times (T - 1)$ matrix representing the horizontal difference of L with element in form of $\mathcal{D}_x(i, j) = L(i, j + 1) - L(i, j)$, and $\mathcal{D}_y(L)$ is an $(N - 1) \times T$ matrix representing the vertical difference of L with element in form of $\mathcal{D}_y(i, j) = L(i + 1, j) - L(i, j)$.

B. Performance Results

We randomly drop entries with data loss rate 10%, 20%, ..., 90%, i.e., with sampling rate 90%, 80%, ..., 10%. Using the remained data as input, we apply the above four schemes to perform data recovery. For the IntelLab, GreenOrbs, and NBDC CTD projects, we test each physical attributes separately. We conduct 10 runs for each case and report an average over these 10 runs.

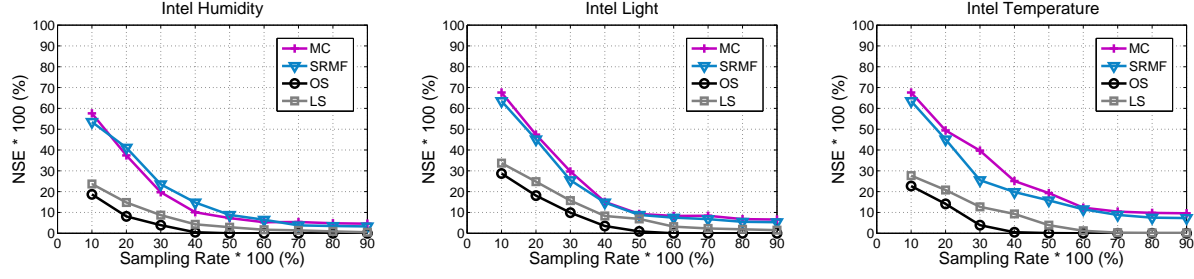


Fig. 6. Comparisons for humidity, light, temperature of the IntelLab project.

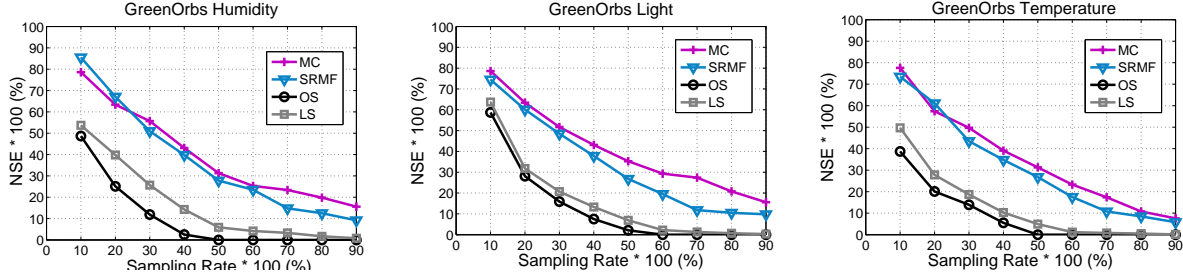


Fig. 7. Comparisons for humidity, light, temperature of the GreenOrbs project.

From Fig. 6, 7, and 8, as the sampling rate increases (or the data loss rate decreases), the recovery errors for all the four algorithms decrease. LS performs much better than MC and SRMF which do not take anomaly readings into consideration. Generally, our scheme achieves recovery error $\leq 5\%$ for sampling rate $\geq 50\%$ and almost exact recovery for sampling rate $\geq 60\%$, while MC and SRMF have error $10\% \sim 15\%$ at sampling rate 90% . Among all projects, the recovery errors of LS for IntelLab are lower than the other two, this is because the data matrices of IntelLab have lower rank. More accurately, lower ratio of rank to matrix dimension, as shown in Fig. 5.

Fig. 6 shows the recovery performance for the IntelLab project. For sampling rate 20% , 30% , 40% , and 50% , SRMF is worse than MC, since SRMF tries to fit the anomaly readings into a smooth surface.

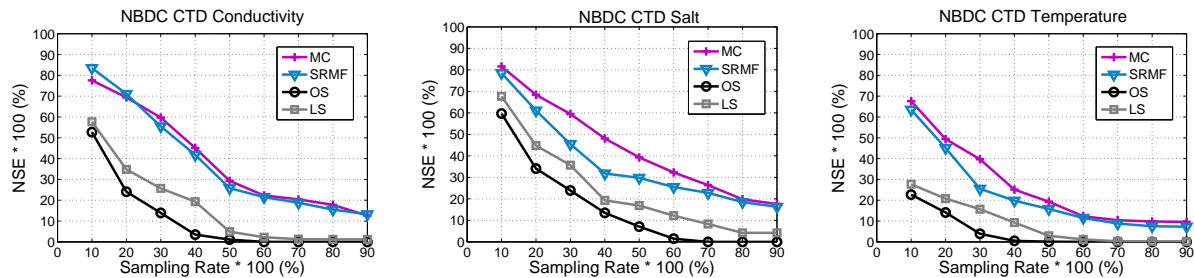


Fig. 8. Comparisons for conductivity, salt, temperature of the NBDC CTD project.

The recovery errors of MC and SRMF do not decrease further with more samples when sampling rate is larger than 60% , LS and OS converge to zero recovery error since LS and OS take anomaly readings into consideration. Among the three physical conditions (humidity, light and temperature), the recovery error for humidity is the lowest for each case, because the humidity matrix has both lowest rank, as shown in Fig. 5, and lowest portion of anomaly readings, as shown in Fig. 4. Although the temperature matrix has relatively low rank too, but it also contains more anomaly readings.

In Fig. 7 and 8, similar results hold for the GreenOrbs and NBDC CTD projects. Compared with IntelLab data matrices, we observe bigger gaps between MC/SRMF and LS/OS. The reasons are: (1) GreenOrbs and NBDC CTD have lower ratios of rank to matrix dimension, as shown in Fig. 5; (2) the dimensions, i.e., $\min(N, T)$, of data matrices in GreenOrbs and NBDC CTD (i.e., 326 and 216, respectively), are much bigger than those in Intellab (i.e., 52), as shown in Table I. For larger matrices, anomaly readings have weak influences. Note that the temperature of NBDC CTD has much lower recovery error than conductivity and salt, the reasons are: (1) the rank of the temperature matrix is much lower, as shown in Fig. 5, and (2) the portion of anomaly readings is also lower.

IX. CONCLUSION

In this paper, we investigated the data recovery problem with considerations of massive missing entries, measurement noise, and anomaly readings. An LS-decomposition approach was proposed, which decomposed a given partially observed corrupted data matrix into a low-rank matrix and a sparse matrix. An accelerated proximal algorithm was devised to solve this problem. Theoretical results were given to guarantee the optimality and the converge rate. Our scheme is more robust than standard matrix completion methods, thus is expected to have better usability.

APPENDIX

A. Proof of Theorem 1

Our proof uses three crucial properties of $\langle \hat{L}, \hat{S} \rangle$:

- Since $\langle L, S \rangle$ is a feasible solution to (10), we have:

$$\|\hat{L}\|_* + \lambda \|\hat{S}\|_1 \leq \|L\|_* + \lambda \|S\|_1. \quad (24)$$

- The triangle inequality implies that:

$$\|\mathcal{P}_O(\hat{L} + \hat{S} - L - S)\|_F \leq \|\mathcal{P}_O(\hat{L} + \hat{S} - R)\|_F + \|\mathcal{P}_O(L + S - R)\|_F \leq 2\delta, \quad (25)$$

since both $\langle \hat{L}, \hat{S} \rangle$ and $\langle L, S \rangle$ are feasible solutions to (10).

- The condition we set in (11) to ensure a unique solution:

$$\|\mathcal{P}_{\mathcal{O}^\perp(\hat{S})}\|_F = \|\mathcal{P}_{\mathcal{O}^\perp(S_0)}\|_F = 0. \quad (26)$$

The first two properties imply that $\hat{X} = \langle \hat{L}, \hat{S} \rangle$ is close to $X = \langle L, S \rangle$. Set $\hat{X} = X + H$, where $H = \langle H_L, H_S \rangle$ and write $H^\Gamma = \mathcal{P}_\Gamma(H)$, $H^{\Gamma^\perp} = \mathcal{P}_{\Gamma^\perp}(H)$ for short. Our aim is to bound $\|H\|_F$, which can be expressed as:

$$\|H\|_F^2 = \|H^\Gamma\|_F^2 + \|H^{\Gamma^\perp}\|_F^2 = \|H^\Gamma\|_F^2 + \|(\mathcal{P}_\mathcal{T} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp})\|_F^2 + \|(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})\|_F^2. \quad (27)$$

It suffices to bound each term in the right-hand-side of (27).

A. Bound the first term of (27).

$$\begin{aligned} \|H^\Gamma\|_F^2 &= \|(H_L + H_S)/2\|_F^2 + \|(H_L + H_S)/2\|_F^2 = \frac{1}{2}\|H_L + H_S\|_F^2 \\ &= \frac{1}{2} (\|\mathcal{P}_\mathcal{O}(H_L + H_S)\|_F^2 + \|\mathcal{P}_{\mathcal{O}^c}(H_L + H_S)\|_F^2) \\ &= \frac{1}{2} (\|\mathcal{P}_\mathcal{O}(H_L + H_S)\|_F^2 + \|\mathcal{P}_{\mathcal{O}^c}(H_L)\|_F^2) \end{aligned} \quad (28)$$

where the last equation uses a condition derived from (26), i.e., $\|\mathcal{P}_{\mathcal{O}^c}(H_S)\|_F^2 = 0$. For the first term of (28), we already have $\|\mathcal{P}_\mathcal{O}(H_L + H_S)\|_F^2 \leq 4\delta^2$ from (25). Then, we bound the second term of (28):

$$\|\mathcal{P}_{\mathcal{O}^c}(H_L)\|_F^2 = \|\mathcal{P}_\mathcal{T}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_F^2 + \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_F^2, \quad (29)$$

and it suffices to bound each term in the right-hand-side.

We start with the second term of (29). Let W be a dual certificate as in [20–22]. Then, $\Lambda = UV^* + W$ obeys $\|\mathcal{P}_{\mathcal{T}^\perp}(\Lambda)\| \leq 1/2$ and $\|\mathcal{P}_{\Omega^\perp}(\Lambda)\|_\infty \leq \lambda/2$. We have:

$$\|L + H_L\|_* \geq \|L + \mathcal{P}_{\mathcal{O}^c}(H_L)\|_* - \|\mathcal{P}_\mathcal{O}(H_L)\|_* \quad (30)$$

and ([26])

$$\|L + \mathcal{P}_{\mathcal{O}^c}(H_L)\|_* \geq \|L\|_* + (1 - \|\mathcal{P}_{\mathcal{T}^\perp}(\Lambda)\|) \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_*. \quad (31)$$

Therefore, with $\|\mathcal{P}_{\mathcal{T}^\perp}(\Lambda)\| \leq 1/2$ and $\|L + H_L\|_* \leq \|L\|_*$, combining (30) and (31) we have:

$$\|L\|_* \geq \|L\|_* + \frac{1}{2} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_* - \|\mathcal{P}_\mathcal{O}(H_L)\|_* \quad (32)$$

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_* \leq 2\|\mathcal{P}_\mathcal{O}(H_L)\|_*. \quad (33)$$

Since the nuclear norm dominated the Frobenius norm $\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_F \leq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_*$, we have

$$\|\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{P}_{\mathcal{O}^c}(H_L))\|_F \leq 2\|\mathcal{P}_\mathcal{O}(H_L)\|_* \leq 2\sqrt{n}\|\mathcal{P}_\mathcal{O}(H_L)\|_F \quad (34)$$

where the second inequality follows from the Cauchy-Schwarz inequality. We know that

$$\|\mathcal{P}_O(H_L)\|_F \leq \|\mathcal{P}_O(H_L + H_S)\|_F \leq 2\delta, \quad (35)$$

therefore,

$$\|\mathcal{P}_{T^\perp}(\mathcal{P}_{O^c}(H_L))\|_F \leq 4\sqrt{n}\delta. \quad (36)$$

We then bound $\|\mathcal{P}_T(\mathcal{P}_{O^c}(H_L))\|_F$ in the first term of (29). Observe that the assumption $\mathcal{P}_T\mathcal{P}_O\mathcal{P}_T \succeq (p/2)\mathcal{I}$ together with $\mathcal{P}_T^2 = \mathcal{P}_T, \mathcal{P}_O^2 = \mathcal{P}_O$, gives:

$$\begin{aligned} \|\mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L))\|_F^2 &= \langle \mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L)), \mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L)) \rangle \\ &= \langle \mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L)), \mathcal{P}_T(\mathcal{P}_{O^c}(H_L)) \rangle \\ &\geq \frac{p}{2} \|\mathcal{P}_T(\mathcal{P}_{O^c}(H_L))\|_F^2. \end{aligned} \quad (37)$$

But since

$$\mathcal{P}_O(\mathcal{P}_O) = 0 = \mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L)) + \mathcal{P}_O\mathcal{P}_{T^\perp}(\mathcal{P}_{O^c}(H_L)), \quad (38)$$

we have

$$\begin{aligned} \|\mathcal{P}_O\mathcal{P}_T(\mathcal{P}_{O^c}(H_L))\|_F &= \|\mathcal{P}_O\mathcal{P}_{T^\perp}(\mathcal{P}_{O^c}(H_L))\|_F \\ &\leq \|\mathcal{P}_{T^\perp}(\mathcal{P}_{O^c}(H_L))\|_F. \end{aligned} \quad (39)$$

Hence, (37) and (39) together give

$$\|\mathcal{P}_T(\mathcal{P}_{O^c}(H_L))\|_F^2 \leq \frac{2}{p} \|\mathcal{P}_{T^\perp}(\mathcal{P}_{O^c}(H_L))\|_F^2. \quad (40)$$

Combining (28)(29)(36)(40), we have:

$$\|H^\Gamma\|_F^2 \leq 2\delta^2 + \frac{2+p}{p} 8n\delta^2. \quad (41)$$

B. Bound the third term of (27). For any matrix pair $X = \langle L, S \rangle$, we define $\|X\|_\diamond = \|L\|_* + \lambda\|S\|_1$.

We have

$$\begin{aligned} \|X + H\|_\diamond &\geq \|X + H^{\Gamma^\perp}\|_\diamond - \|H^\Gamma\|_\diamond, \\ \|X + H^{\Gamma^\perp}\|_\diamond &\geq \|X\|_\diamond + (3/4 - \|\mathcal{P}_{T^\perp}(\Lambda)\|) \|\mathcal{P}_{T^\perp}(H_L^{\Gamma^\perp})\|_* + (3\lambda/4 - \|\mathcal{P}_{\Omega^\perp}(\Lambda)\|_\infty) \|\mathcal{P}_{\Omega^\perp}(H_S^{\Gamma^\perp})\|_* \\ &\geq \|X\|_\diamond + \frac{1}{4} (\|\mathcal{P}_{T^\perp}(H_L^{\Gamma^\perp})\|_* + \lambda \|\mathcal{P}_{\Omega^\perp}(H_S^{\Gamma^\perp})\|_*), \end{aligned}$$

where the second inequality follows from Lemma 5 of [22].

Therefore, we have

$$\|\mathcal{P}_{T^\perp}(H_L^{\Gamma^\perp})\|_* + \lambda \|\mathcal{P}_{\Omega^\perp}(H_S^{\Gamma^\perp})\|_* \leq 4\|H^\Gamma\|_\diamond. \quad (42)$$

For any matrix $Y \in \mathbb{R}^{n \times n}$, we have the following inequalities:

$$\|Y\|_F \leq \|Y\|_* \leq \sqrt{n}\|Y\|_F, \quad \frac{1}{\sqrt{n}}\|Y\|_F \leq \lambda\|Y\|_1 \leq \sqrt{n}\|Y\|_F, \quad (43)$$

where we assume $\lambda = \frac{1}{\sqrt{n}}$. Then,

$$\begin{aligned}
\|(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})\|_F^2 &\leq \|\mathcal{P}_{\mathcal{T}^\perp}(H_L^{\Gamma^\perp})\|_F + \|\mathcal{P}_{\Omega^\perp}(H_S^{\Gamma^\perp})\|_F \\
&\leq \|\mathcal{P}_{\mathcal{T}^\perp}(H_L^{\Gamma^\perp})\|_* + \lambda\sqrt{n}\|\mathcal{P}_{\Omega^\perp}(H_S^{\Gamma^\perp})\|_1 \\
&\leq 4\sqrt{n}\|H^\Gamma\|_\diamond = 4\sqrt{n}(\|H_L^\Gamma\|_* + \|H_S^\Gamma\|_1) \\
&\leq 4n(\|H_L^\Gamma\|_F + \|H_S^\Gamma\|_F) = 4\sqrt{2}n\|H^\Gamma\|_F
\end{aligned} \tag{44}$$

C. Bound the second term of (27). By Lemma 6 of [22], we have

$$\|\mathcal{P}_\Gamma(\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp})\|_F^2 \geq \frac{1}{4}\|(\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp})\|_F^2. \tag{45}$$

But since $\mathcal{P}_\Gamma(H^{\Gamma^\perp}) = 0 = \mathcal{P}_\Gamma(\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp}) + \mathcal{P}_\Gamma(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})$, we have

$$\begin{aligned}
\|\mathcal{P}_\Gamma(\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp})\|_F &= \|\mathcal{P}_\Gamma(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})\|_F \\
&\leq \|(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})\|_F.
\end{aligned} \tag{46}$$

Combing (45) and (46), we have

$$\|(\mathcal{P}_{\mathcal{T}} \times \mathcal{P}_\Omega)(H^{\Gamma^\perp})\|_F^2 \leq 4\|(\mathcal{P}_{\mathcal{T}^\perp} \times \mathcal{P}_{\Omega^\perp})(H^{\Gamma^\perp})\|_F^2, \tag{47}$$

together with (27)(41)(44), leads us to the final result:

$$\begin{aligned}
\|H\|_F^2 &\leq 5 \times 16 \times 2n^2\|H^\Gamma\|_F^2 + \|H^\Gamma\|_F^2 \\
&= 320n^2\delta^2(8n + 8n/p + 1) + \frac{2+p}{p}8n\delta^2 + 2\delta^2.
\end{aligned} \tag{48}$$

Therefore, we obtain the desired result,

$$\|H\|_F \leq (8n\sqrt{40n + 40n/p + 5} + \sqrt{2})\delta. \tag{49}$$

B. Proof of Theorem 2

First, we need the following three key inequalities.

Lemma 3. Let $\{a_k, b_k\}$ be positive sequences satisfying

$$a_k - a_{k+1} \geq b_{k+1} - b_k, \quad \forall k \leq 1, \text{ with } a_1 + b_1 \leq c, \quad c > 0. \tag{50}$$

Then, $a_k \leq c$ for every $k \geq 1$.

By deduction, we can easily have:

Lemma 4. The positive sequence $\{t_k\}$ generated in our algorithm via $t_{k+1} \leftarrow \frac{1+\sqrt{4t_k^2+1}}{2}$ with $t_{-1} = 1$ satisfies $t_k \geq (k+1)/2$ for all $k \geq 1$.

Let $v_k \triangleq F(X_k) - F(X^*)$, $u_k \triangleq t_k X_k - (t_k - 1)X_{k_1} - X^*$. Similarly as is proved in [28], we have:

Lemma 5. *The sequence $\{X_k\}$ generated in our algorithm satisfies:*

$$t_k^2 v_k - t_{k+1}^2 v_{k+1} \geq \|u_{k+1}\|_F^2 - \|u_k\|_F^2. \quad (51)$$

Now, we are ready to prove our theorem. Let $a_k \triangleq t_k^2 v_k$, $b_k \triangleq \|u_k\|_F^2$, $c \triangleq \|X_{k_0} - X^*\|_F^2$, then from Lemma 4,

$$a_k - a_{k+1} \geq b_{k+1} - b_k. \quad (52)$$

Assume that $a_k + b_k \leq c$ holds true for $k > 1$, combining Lemma 50, we obtain that:

$$t_k^2 v_k \leq \|X_k - X^*\|_F^2, \quad (53)$$

which combined with Lemma 4 yields

$$F(X_k) - F(X^*) \leq \frac{4\|X_1 - X^*\|_F^2}{k^2}. \quad (54)$$

C. Proof of Lemma 2

From Theorem 2, we know that:

$$\lim_{k \rightarrow \infty} F(X_k) - F(X^*) \rightarrow 0, \quad (55)$$

which means that our algorithm converges to its globe optimal for large enough k .

For any $\varepsilon > 0$, to guarantee $F(X_k) - F(X^*) \leq \varepsilon$, then:

$$\frac{4\|X_1 - X^*\|_F^2}{k^2} \leq \varepsilon, \quad (56)$$

$$k \geq \frac{4\|X_1 - X^*\|_F^2}{\sqrt{\varepsilon}} = O(1/\sqrt{\varepsilon}). \quad (57)$$

REFERENCES

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Elsevier Computer Networks*, 38(4), pp. 393-422, 2002.
- [2] Y. Wang, W. Fu, and D.P. Agrawal, "Gaussian versus uniform distribution for intrusion detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 24(2), pp. 342-355, 2013.
- [3] L. He, P. Cheng, Y. Gu, J. Pan, T. Zhu, and Cong Liu, "Mobile-to-mobile energy replenishment in mission-critical robotics sensor networks," *IEEE INFOCOM*, 2014.

- [4] X. Liu, J. Cao, S. Tang, and P. Guo, "A generalized coverage-preserving scheduling in WSNs: a case study in structural health monitoring," *IEEE INFOCOM*, 2014.
- [5] R.G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, 331(6081), pp. 717-719, 2011.
- [6] L. Kong, M. Xia, X.-Y. Liu, M. -Y. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," *IEEE INFOCOM*, 2013.
- [7] G. Chen, X.-Y. Liu, L. Kong, J.-L. Lu, Y. Gu, W. Shu, and M.-Y. Wu, "Multiple Attributes-based Data Recovery in Wireless Sensor Network," *IEEE Globecom*, 2013.
- [8] Y. Yang, J. Ma, and S. Osher, "Seismic data reconstruction via matrix completion," *UCLA CAM Report*, pp. 12-14, 2012.
- [9] X. Liu, J. Luo, C. Deng, A.V. Vasilakos, "DECA: Recovering fields of physical quantities from incomplete sensory data," *IEEE SECON*, pp. 182-190, 2012.
- [10] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks," *IEEE Transactions on Wireless Communications*, Vol. 12, No. 2, pp. 850-861, 2013.
- [11] K. Xie, L. Wang, X. Wang, J. Wen, and G. Xie, "Learning from the past: intelligent on-line weather monitoring based on matrix completion", *IEEE ICDCS*, 2014.
- [12] E.J. Candès and B.Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, 9(6), pp. 717-772, 2009.
- [13] E.J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, Vol. 56, No. 5, pp. 2053-2080, 2010.
- [14] S. Becker, E.J. Candès and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Mathematical Programming Computation*, Vol. 3, No. 3, 2011.
- [15] IntelLab data. <http://www.select.cs.cmu.edu/data/labapp3/index.html>.
- [16] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, and X.-Y. Li, "Does wireless sensor network scale? a measurement study on greenorbs," *IEEE INFOCOM*, pp. 873-881, 2011.
- [17] NBDC-CTD data. http://tao.noaa.gov/refreshed/ctd_delivery.php.
- [18] M.C. Vuran, O.B. Akan, and I.F. Akyildiz, "Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks," *Elsevier Computer Networks*, 45(3), pp. 245-261, 2004.
- [19] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks", *IEEE Journal Selected Areas in Communications (JSAC)*, 29(8), pp.

- 1660-1670, 2011.
- [20] E.J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM (JACM)*, 58(3): 11, 2011.
 - [21] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, “Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization,” *Proc. of Neural Information Processing Systems (NIPS)*, 2009.
 - [22] Z. Zhou, X. Li, J. Wright, E.J. Candès, and Y. Ma, “Stable principal component pursuit,” *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 1518-1522, 2010.
 - [23] Y.-C. Chen, L. Qiu, Y. Zhang, G. Xue, and Z. Hu, “Robust network compressive sensing,” *ACM MobiCom*, 2014.
 - [24] E.J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, 52(2), pp. 489-509, 2006.
 - [25] E.J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Transactions on Information Theory*, 52(12), pp. 5406-5425, 2006.
 - [26] E.J. Candès and Y. Plan, “Matrix completion with noise,” *Proceedings of the IEEE*, 98(6): 925-936, 2010.
 - [27] N. Parikh and S. Boyd, “Proximal Algorithms,” *Foundations and Trends in Optimization*, 1(3), pp. 123-231, 2014.
 - [28] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, Vol. 3, No. 2, pp. 183-202, 2009.
 - [29] R. Meka, P. Jain, C. Caramanis, and I. S. Dhillon, “Rank minimization via online learning,” *ICML*, pp. 656-663, 2008.